



การวิเคราะห์ข้อมูลจำแนกประเภทของระดับการเป็นโรคหัวใจ ด้วยตัวแบบเชิงเส้นนัยทั่วไปและอัตราส่วนออดส์

A Categorical Data Analysis of Heart Disease Levels Using Generalized Linear Models and Odds Ratios

วริฐา เกตุแก้ว¹ สุพัตรา เอี้ยงอารีย์ และ วีรานันท์ พงศาภักดี^{1*}

บทคัดย่อ

การวิเคราะห์ข้อมูลจำแนกประเภทเพื่อศึกษาปัจจัยเสี่ยงที่ส่งผลต่อความน่าจะเป็นของระดับการเป็นโรคหัวใจ พิจารณาจากตัวแบบเชิงเส้นนัยทั่วไปต่าง ๆ ที่เหมาะสมกับข้อมูลโดยมีตัวแปรตอบสนองสถานะภาพของ คนไข้โรคหัวใจ ซึ่งแบ่งออกเป็น 2 กลุ่มคือ การเสียชีวิตอยู่และการมีชีวิตรอดอยู่ และสถานะภาพที่ได้รับการวินิจฉัยโรคหัวใจ ซึ่งแบ่งออกเป็น 4 กลุ่มคือ การไม่เป็นโรคหัวใจและยังมีชีวิต การเป็นโรคหัวใจและยังมีชีวิต การไม่เป็นโรคหัวใจแต่เสียชีวิต และการเป็นโรคหัวใจแต่เสียชีวิต ส่วนตัวแปรอธิบายหรือปัจจัยเสี่ยงด้านการแพทย์ที่ศึกษา คือ อายุ ดัชนีมวลกาย ระดับความดันโลหิต ระดับคอเลสเตอรอลในเลือด ระดับความเครียด และการมีผลวินิจฉัยของโรคหัวใจแบบขาดเลือด ข้อมูลเป็นข้อมูลจริง (Afifi and Azen, 1979) จำนวน 200 คน วิเคราะห์หาค่าด้วยตัวแบบ โพรบิต ตัวแบบลอจิสต์ และตัวแบบลอจิสต์เทียบกลุ่มฐาน โดยใช้โปรแกรม SAS Enterprise Guide 5.1 ผลการวิจัยพบว่า ปัจจัยเสี่ยงที่มีผลต่อการเสียชีวิตเนื่องด้วยโรคหัวใจภายใต้ตัวแบบโพรบิตและตัวแบบลอจิสต์ คือ อายุ ระดับความดันโลหิต ระดับคอเลสเตอรอลในเลือด และการมีผลวินิจฉัยของโรคหัวใจแบบขาดเลือดอย่างมีนัยสำคัญทางสถิติที่ 0.05 ส่วนการวิเคราะห์ข้อมูลด้วยตัวแบบลอจิสต์เทียบกลุ่มฐานคือ อายุ และระดับความดันโลหิต โดยตัวแบบเชิงเส้นนัยทั่วไปทั้ง 3 ตัวแบบ พบว่ามีภาวะสารูปดีและเหมาะสมกับข้อมูลอย่างมีนัยสำคัญทางสถิติที่ 0.05 อัตราส่วนออดส์ (odds ratio) เช่นสำหรับปัจจัยอายุ พบว่า ความน่าจะเป็นของระดับการเป็นโรคหัวใจที่จะเสี่ยงต่อการเสียชีวิตของคนไข้กลุ่มที่มีอายุมากจะเป็น $\exp(1.1671) = 3.21266$ เท่า ของคนไข้กลุ่มที่มีอายุน้อยกว่าที่เป็นกลุ่มฐาน เมื่อปัจจัยอื่น ๆ คงตัว ภายใต้ตัวแบบลอจิสต์

¹ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์ นครปฐม 73000

*Corresponding Author, E-mail: pongsapakdee_v@silpakorn.edu, veeranun@hotmail.com

ABSTRACT

The analysis of risk factors affecting the probability of category levels due to heart disease is evaluated. Two cases are considered, case 1: responses are divided into two groups: death and survivors, case 2: responses are divided into 4 groups: non-heart disease and living, heart disease but still living, non-heart disease but death, and heart disease and death. The medical risk factors consist of age, body mass index, blood pressure levels, cholesterol levels, stress levels. The research real data with 200 patients were from Afifi and Azen (1979). All analyses are performed using the probit model, logit model and baseline-category logit model run with SAS Enterprise Guide 5.1. The results show the risk factors under the probit model and the logit model are age, blood pressure levels, cholesterol levels and the diagnosis of cardiac ischemia, significantly at 0.05. Those factors from the baseline-category logit models are also age and blood pressure levels, significantly at 0.05. The three generalized linear models are all statistically adequate of fits for the data, significantly at 0.05. It is also found from odds ratio such that the risk factors for age is 3.21266 under the logit model.

คำสำคัญ: ตัวแบบโพรบิต ตัวแบบลอจิต ตัวแบบลอจิตเทียบกลุ่มฐาน ปัจจัยเสี่ยงด้านการแพทย์ โรคหัวใจ

Keyword: Probit model, Logit model, Baseline-category logit model, Medical risk factors, Heart disease

บทนำ (Introduction)

มนุษย์จะดำรงอยู่ได้ด้วยการทำงานของอวัยวะต่าง ๆ ทั้งภายในและภายนอกร่างกาย การแบ่งส่วนประกอบของร่างกายออกเป็นระบบต่าง ๆ ซึ่งต้องพึ่งพาและทำงานสัมพันธ์กันช่วยให้มีการศึกษาดูแลและการรักษาที่ชัดเจนยิ่งขึ้น เช่น ระบบย่อยอาหาร ระบบขับถ่าย ระบบประสาท ระบบหัวใจ และการไหลเวียนเลือด ฯลฯ ทุกระบบมีความสำคัญต่อร่างกายมนุษย์ทั้งสิ้น เราต้องรักษาสุขภาพร่างกายให้สมบูรณ์เพื่อการดำรงประสิทธิภาพการทำงานของระบบต่าง ๆ ในร่างกายให้อยู่ในภาวะปกติและให้สามารถดำเนินกิจกรรมในชีวิตประจำวันได้อย่างมีประสิทธิภาพ

ระบบต่าง ๆ ในร่างกายประกอบด้วยระบบหลายระบบ ในที่นี้จะกล่าวถึงระบบหัวใจและการไหลเวียนโลหิตซึ่งประกอบด้วยเลือด หลอดเลือดและหัวใจ โดยที่กลุ่มเซลล์มีการก่อสร้างหัวใจขึ้นตั้งแต่ในครรภ์มารดา หัวใจเริ่มต้นไม่มีวันหยุดตราบเท่าที่ยังมีชีวิต หัวใจไม่มีวันหยุดพักจึงเป็นอวัยวะที่มีความมหัศจรรย์และแข็งแรงมาก หัวใจมีหน้าที่สูบฉีดเลือดให้ไหลไปตามหลอดเลือดไปยังส่วนต่าง ๆ ของร่างกายแล้วไหลกลับคืนสู่หัวใจโดยอาศัยโครงสร้างของกล้ามเนื้อหัวใจและระบบนำไฟฟ้าภายในหัวใจซึ่งสร้างและควบคุมจังหวะการเต้นของหัวใจ โดยปกติหัวใจของคนที่มีสุขภาพสมบูรณ์มีขนาดประมาณกำปั้นของเจ้าของ การทำงานของหัวใจจะอาศัยกล้ามเนื้อหัวใจสามารถปล่อยกระแสไฟฟ้าได้เอง โดยมีอัตราการปล่อยไฟฟ้าประมาณ 60-100 ครั้งต่อนาทีจะกระจายออกไปตาม

เซลล์นำไฟฟ้า (สมาคมแพทย์โรคหัวใจแห่งประเทศไทยในพระบรมราชูปถัมภ์, <http://www.thaiheart.org>) ดังนั้นหัวใจซึ่งมีหน้าที่สำคัญควรได้รับการดูแล และหลีกเลี่ยงปัจจัยเสี่ยงที่อาจส่งผลกระทบต่อการศึกษาโรคหัวใจได้

ความเป็นจริงคำว่า “โรคหัวใจ” มีความหมายกว้างมาก เป็นโรคที่พบบ่อยมากสำหรับผู้ป่วยและพบบ่อยโดยเฉพาะโรคหัวใจที่เกิดกับหลอดเลือดที่ไปเลี้ยงหัวใจ นอกจากนั้นยังเกิดโรคหัวใจที่กล้ามเนื้อหัวใจ ลิ้นหัวใจ เยื่อหุ้มหัวใจ เนื่องจากปัญหาโรคหัวใจเป็นปัญหาที่มีผลกระทบต่อสุขภาพมนุษย์ในอันดับต้น ๆ โรคหนึ่ง ทั่วโลกมีจำนวนคนที่เป็นโรคหัวใจเพิ่มขึ้นในทุก ๆ ประเทศ อาจเนื่องด้วยอิทธิพลของโลกไร้พรมแดนในปัจจุบัน วิธีการดำเนินชีวิตของคนเราเปลี่ยนแปลงไปจากเดิม มีการทำงานที่หนักขึ้น ชีวิตมีการแข่งขันอยู่ตลอดเวลา การดูแลสุขภาพตัวเองกลับลดลง การออกกำลังกายไม่สม่ำเสมอ ซึ่งอาจเกิดจากสาเหตุในภาพรวมเหล่านี้ที่ทำให้มีโรคภัยไข้เจ็บมาเบียดเบียนชีวิตของคน

กระทรวงสาธารณสุขเผยคนทั่วโลกเสียชีวิตจากโรคหัวใจและหลอดเลือดเป็นอันดับ 1 ของโลกมาก และกล่าวถึงสถานการณ์ผู้ป่วยด้วยโรคหัวใจและหลอดเลือด ในงานวันหัวใจโลก ปี 2553 ว่าคนทั่วโลกเสียชีวิตด้วยโรคหัวใจและหลอดเลือดกว่า 17.5 ล้านคน ซึ่งถือเป็นสาเหตุการเสียชีวิตอันดับ 1 ของทั่วโลก และมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง โดยองค์การอนามัยโลกคาดการณ์ว่า ในปี 2558 จะมีผู้เสียชีวิตทั่วโลกจากโรคนี้ถึง 20 ล้านคน (<http://www.wikipedia.org>)

ในปัจจุบันโรคหัวใจเป็นสาเหตุของการป่วยและเสียชีวิตอันดับต้น ๆ ที่คร่าชีวิตพลเมืองโลกมากที่สุด จากสถิติพบว่าในทุก ๆ 2 วินาที จะมีคนเสียชีวิตด้วยโรคนี้ 1 คน ผู้ร้ายตัวจริงเบื้องหลังก็คือการที่เส้นเลือดแดงเกิดภาวะแข็งตัวส่งผลให้เกิดการอุดตันในเส้นเลือดซึ่งเป็นสาเหตุสำคัญ ปัจจัยเสี่ยงที่สำคัญนอกเหนือจากปัจจัยด้านพันธุกรรมอาจเป็น อายุและเพศ ผู้เสียชีวิตส่วนใหญ่จะมีอายุตั้งแต่ 55 ปีขึ้นไป ยังมีปัจจัยอื่น ๆ เช่น ระดับคอเลสเตอรอลมากเกินไปในหลอดเลือด การสูบบุหรี่ การดื่มแอลกอฮอล์ การใชยาคุมกำเนิด การมีโรคความดันโลหิตสูง ภาวะอ้วน โรคเบาหวาน ไตรกลีเซอไรด์สูงและความเครียด (<http://www.saimhelath.net>) ซึ่งน่าสนใจศึกษาว่าจะส่งผลกระทบในเชิงสถิติต่อการเป็นโรคหัวใจอย่างไร

ดังนั้นปัญหาของโรคหัวใจในปัจจุบันนับวันยิ่งมีโอกาสมากขึ้น และยังไม่สามารถแก้ไขหรือป้องกันปัญหาให้หมดไปได้ ซึ่งการเป็นโรคหัวใจอาจเป็นผลมาจากปัจจัยเสี่ยงต่าง ๆ มากมาย โดยปัจจัยเสี่ยงหลายปัจจัยอาจควบคุมหรือหลีกเลี่ยงได้โดยง่าย แต่อาจมีบางปัจจัยที่อาจส่งผลกระทบต่อการเป็นโรคหัวใจโดยตรงหรืออาจเสี่ยงต่อการเสียชีวิต ถ้าไม่มีการศึกษาหรือการป้องกันแต่เนิ่น ๆ

ในการวิจัยครั้งนี้ผู้วิจัยสนใจการวิเคราะห์ปัจจัยที่มีผลต่อการเสียชีวิตเนื่องด้วยโรคหัวใจ โดยอาศัยการวิเคราะห์ข้อมูลจำแนกประเภทด้วยตัวแบบเชิงเส้นทั่วๆไป (generalized linear models; GLMs) (Nelder and Wedderburn, 1972, Agresti, 2002, วีรานันท์, 2555) ที่สอดคล้องกับข้อมูลโดยใช้ตัวแบบ 3 ตัวแบบ โดยมีขอบเขตการวิจัยดังนี้

1. ตัวแบบโพรบิต (probit model) (Bliss, 1935) ซึ่งนิยมใช้สำหรับตัวแปรอธิบายแบบต่อเนื่องอย่างน้อย 1 ตัวแปร
2. ตัวแบบลอจิสต์ (logit model) (Agresti, 2002) ซึ่งนิยมใช้สำหรับตัวแปรอธิบายแบบจำแนกประเภทหรือเชิงกลุ่มทั้งหมด แต่ถ้าสนใจศึกษาตัวแปรอธิบายแบบต่อเนื่องด้วยบางตัวแปร จะเรียกตัวแบบนี้ใหม่ว่า ตัวแบบ

การถดถอยลอจิสติก (logistic regression model) (Agresti, 2002) แทน ซึ่งในการวิจัยนี้ไม่สนใจตัวแบบหลังนี้ เนื่องจากอาจมีปัญหาที่ต้องจัดแบ่งตัวแปรต่อเนื่องทุกตัวออกเป็น 10 กลุ่ม เพื่อให้ตัวแบบมีความถูกต้องและเชื่อถือได้ (Hosmer and Lemeshow, 1989)

3. ตัวแบบลอจิสติกเทียบกับกลุ่มฐาน (baseline-category logit model) (Agresti, 2002) ซึ่งนิยมใช้สำหรับตัวแปรอธิบายแบบจำแนกประเภทหรือเชิงกลุ่มหรืออาจผสมตัวแปรต่อเนื่องบ้าง ซึ่งเป็นตัวแบบเชิงสถิติขั้นสูงที่มีประสิทธิภาพดี และนิยมนำมาประยุกต์ด้านการแพทย์ เพื่อศึกษาปัจจัยเสี่ยงและการพยากรณ์กลุ่มที่สนใจได้หลายกลุ่ม เช่น ระดับความรุนแรงของโรค ประเภทหรืออาการของโรค (จุฬามาตและคณะ, 2557)

วัตถุประสงค์และขอบเขตของงานวิจัยนี้ เพื่อศึกษาโครงสร้างตัวแบบ GLMs ใหม่จำนวน 3 ตัวแบบ ข้างต้น ทดสอบความเหมาะสมของทุกตัวแบบ และสืบสวนหาสารสนเทศจากอิทธิพลหลักต่าง ๆ (main effects) ที่ส่งผลกระทบในเชิงสถิติอย่างมีนัยสำคัญเชิงสถิติ โดยยังไม่รวมการศึกษาอิทธิพลร่วม (interaction effects) เนื่องจากอิทธิพลร่วมในตัวแบบ GLMs ควรเป็นอิทธิพลร่วมระหว่างปัจจัยจำแนกประเภทหรือเชิงกลุ่มปัจจัยตั้งแต่ 2 ปัจจัยขึ้นไปเพื่อให้สามารถตีความหมายของผลลัพธ์ได้ (Agresti, 2002; 2013) งานวิจัยนี้จึงศึกษาลักษณะตัวแปรที่เหมาะสมตามสมบัติของแต่ละตัวแบบ ซึ่งบางตัวแบบเหมาะที่จะใช้ตัวแปรต่อเนื่องและบางตัวแบบเหมาะที่จะใช้ตัวแปรจำแนกประเภทหรือเชิงกลุ่ม และในทางทฤษฎีการศึกษาอิทธิพลร่วมระหว่างตัวแปรต่อเนื่องหรือระหว่างตัวแปรต่อเนื่องกับตัวแปรเชิงกลุ่มนั้นไม่มีความหมายและไม่สามารถตีความหมายของผลลัพธ์ได้ (Agresti, 2013) ดังนั้นเพื่อให้มีความคงเส้นคงวาของทุกตัวแบบของการวิจัยนี้ จึงเน้นศึกษาอิทธิพลหลัก หรือปัจจัยเสี่ยงหลักเป็นสำคัญ ส่วนการศึกษาอิทธิพลร่วมควรเป็นงานวิจัยครั้งต่อไป ที่ใช้เฉพาะตัวแปรอธิบายจำแนกประเภททุกตัวแบบ และสามารถตีความหมายของผลลัพธ์ของอิทธิพลร่วมระหว่างตัวแปรจำแนกประเภทหรือเชิงกลุ่มของทุกตัวแบบได้ทุกตัวแบบด้วย นอกจากนี้งานวิจัยนี้ยังสนใจการพยากรณ์ความน่าจะเป็นของระดับการเป็นโรคหัวใจด้วยตัวแบบเชิงเส้นน้อยทั่วไปภายใต้ตัวแบบอิทธิพลหลักที่ได้ทดสอบด้วยสถิติดีไวเนียนส์ว่ามีภาวะสารูปดีและเหมาะสมกับข้อมูลจริงที่ศึกษาอย่างมีนัยสำคัญเชิงสถิติ รวมทั้งการสรุปผลด้วยเทอมอัตราส่วนออดส์ ส่วนการเปรียบเทียบตัวแบบเพื่อเลือกตัวแบบที่เหมาะสมที่สุด ควรใช้สำหรับวิธีการทางสถิติเช่นการจำลองแบบ (simulation) และอาศัยผลสรุปจากข้อมูลจำนวนมากหรือหลายชุดแทน

วิธีการดำเนินการวิจัย (Research Methodology)

งานวิจัยนี้สนใจศึกษาการวิเคราะห์ข้อมูลจำแนกประเภทด้วยตัวแบบ GLMs โดยกำหนดตัวแปรตอบสนองที่ใช้ในงานวิจัย แบ่งออกเป็น 3 กรณี ดังนี้

กรณีที่ 1 การวิเคราะห์ข้อมูลโดยใช้ตัวแบบโพรบิต (probit models) เมื่อตัวแปรตอบสนองคือ สถานภาพของคนที่ได้รับการตรวจโรคหัวใจที่มี 2 กลุ่ม คือ 1 แทนเสียชีวิต และ 2 แทนมีชีวิตอยู่

กรณีที่ 2 การวิเคราะห์ข้อมูลโดยใช้ตัวแบบลอจิสติก (logit models) เมื่อตัวแปรตอบสนองคือ สถานภาพของคนที่ได้รับการตรวจโรคหัวใจที่มี 2 กลุ่ม คือ 1 แทนเสียชีวิต และ 2 แทนมีชีวิตอยู่

กรณีที่ 3 การวิเคราะห์ข้อมูลโดยใช้ตัวแบบลอจิสติกเทียบกับกลุ่มฐาน (baseline-category logit models) เมื่อตัวแปรตอบสนอง คือ สถานภาพของคนที่ได้รับการวินิจฉัยโรคหัวใจที่มี 4 กลุ่มแบบนามบัญญัติ (nominal

scale) ดังนี้ 1 แทน การไม่เป็นโรคหัวใจและยังมีชีวิตอยู่, 2 แทน การเป็นโรคหัวใจแต่ยังมีชีวิตอยู่, 3 แทน การไม่เป็นโรคหัวใจแต่เสียชีวิต และ 4 แทน การเป็นโรคหัวใจและเสียชีวิต

การประมวลผลข้อมูลในทุกกรณีอาศัยโปรแกรมสำเร็จรูป SAS (Enterprise Guide 5.1) ลิขสิทธิ์ของภาคิวิศาสติติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร โดยมีตัวแปรอธิบายหรือปัจจัยเสี่ยงที่นำมาศึกษาทั้งหมด 9 ตัวแปร ดังตารางที่ 1

ตารางที่ 1 ตัวแปรอธิบายหรือปัจจัยเสี่ยงที่ใช้ในการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลโดยใช้ตัวแบบพหุคูณ		การวิเคราะห์ข้อมูลโดยใช้ตัวแบบลอจิสติก และตัวแบบลอจิสติกเทียบกับกลุ่มฐาน
ตัวแปรต่อเนื่อง	ตัวแปรจำแนกประเภทหรือเชิงกลุ่ม	ตัวแปรจำแนกประเภทหรือเชิงกลุ่ม
1. อายุ (AGE)	4. ระดับความดันโลหิตครั้งที่ 1 (BP1) มี 2 กลุ่ม: ปกติ และ สูง	1. อายุ (AGE_G) มี 2 กลุ่ม: ต่ำกว่า 55 ปี, ตั้งแต่ 55 ปีขึ้นไป
2. ดัชนีมวลกายครั้งที่ 1 (BMI1)	5. ระดับความดันโลหิตครั้งที่ 2 (BP2) มี 2 กลุ่ม: ปกติ, สูง	2. ดัชนีมวลกายครั้งที่ 1 (BMI1_G) มี 4 กลุ่ม: ต่ำกว่าเกณฑ์, ปกติ, สูงกว่าเกณฑ์, ภาวะโรคอ้วน
3. ดัชนีมวลกายครั้งที่ 2 (BMI2)	6. ระดับคอเลสเตอรอลในเลือดครั้งที่ 1 (CHOL1) มี 3 กลุ่ม: ปกติ, เริ่มสูง, สูง	3. ดัชนีมวลกายครั้งที่ 2 (BMI2_G) มี 4 กลุ่ม: ต่ำกว่าเกณฑ์, ปกติ, สูงกว่าเกณฑ์, ภาวะโรคอ้วน
	7. ระดับคอเลสเตอรอลในเลือดครั้งที่ 2 (CHOL2) มี 3 กลุ่ม: ปกติ, เริ่มสูง, สูง	4. ระดับความดันโลหิตครั้งที่ 1 (BP1) มี 2 กลุ่ม: ปกติ, สูง
	8. ระดับความเครียด (SES) มี 3 กลุ่ม: ปกติ, ปานกลาง, สูง	5. ระดับความดันโลหิตครั้งที่ 2 (BP2) มี 2 กลุ่ม: ปกติ, สูง
	9. การถูกวินิจฉัยของโรคหัวใจขาดเลือด (IHDDX) มี 4 กลุ่ม: ไม่ใช่, กล้ามเนื้อหัวใจตาย, กล้ามเนื้อหัวใจขาดเลือด, อื่น ๆ	6. ระดับคอเลสเตอรอลในเลือดครั้งที่ 1 (CHOL1) มี 3 กลุ่ม: ปกติ, เริ่มสูง, สูง
		7. ระดับคอเลสเตอรอลในเลือดครั้งที่ 2 (CHOL2) มี 3 กลุ่ม: ปกติ, เริ่มสูง, สูง
		8. ระดับความเครียด (SES) มี 3 กลุ่ม: ปกติ, ปานกลาง, สูง
		9. การถูกวินิจฉัยของโรคหัวใจขาดเลือด (IHDDX) มี 4 กลุ่ม: ไม่ใช่, กล้ามเนื้อหัวใจตาย, กล้ามเนื้อหัวใจขาดเลือด, อื่น ๆ

หมายเหตุ: ตัวแปรอธิบายจำแนกประเภทหรือเชิงกลุ่มในตารางที่ 1 แบ่งตามเกณฑ์ด้านการแพทย์ และในการวิเคราะห์จะใช้กลุ่มสุดท้ายเป็นกลุ่มฐานในการเปรียบเทียบตาม Default ใน SAS

การเลือกปัจจัยเสี่ยงหรือตัวแปรเข้าตัวแบบอาศัยวิธีการเลือกตัวแปรแบบขั้นตอนโดยวิธีอัตราส่วนที่น่าจะเป็น (stepwise likelihood ratio: LR) การประมาณค่าพารามิเตอร์ใช้วิธีภาวะน่าจะเป็นสูงสุดแบบวนซ้ำ (iterative maximum likelihood) การวัดความเกี่ยวพันของความเป็นของการพยากรณ์และตอบสนองค่าสังเกตใช้ตัวสถิติสอเมอร์สดี (Somers'D) ตัวสถิติซี (C) ตัวสถิติทาวของเคนดอลล์ (Kendall's Tau) และการทดสอบพารามิเตอร์ของตัวแบบอาศัยตัวสถิติทดสอบวาลด์ (Wald test) ที่ระดับนัยสำคัญ 0.05 ส่วนการทดสอบภาวะสารถบึกของตัวแบบทั้ง 3 กรณีอาศัยตัวสถิติอัตราส่วนภาวะน่าจะเป็น หรือ ดีไวเียนซ์ (likelihood ratio statistic or deviance) ซึ่งมีการแจกแจงประมาณด้วยการแจกแจงไคกำลังสอง เมื่อตัวอย่างมีขนาดใหญ่ ณ องศาอิสระที่สอดคล้องกัน

ผลการวิจัย (Results)

ผลการวิจัยแสดงไว้เป็น 3 กรณีภายใต้ตัวแบบ GLMs แต่ละตัวแบบ ดังนี้

กรณีที่ 1 การวิเคราะห์ข้อมูลของตัวแปรตอบสนอง ภายใต้ตัวแบบโพรบิต

ตารางที่ 2 ความเกี่ยวพันของตัวแปรตอบสนองกับปัจจัยเสี่ยง ตัวแบบโพรบิต

ปัจจัยเสี่ยง (risk factors)	DF	Wald	P-value
AGE	1	27.0733	<0.0001
BP1	1	4.8269	0.0280
CHOL2	2	13.6248	0.0364
IHDDX	3	11.7021	0.0085

จากตารางที่ 2 พบว่าตัวแปรอธิบายที่มีความเกี่ยวพันกับตัวแปรตอบสนอง ตัวแปรที่ถูกคัดเลือกเข้าตัวแบบโพรบิต ด้วยวิธีการตัวแบบขั้นตอนได้แก่ อายุ (AGE), ระดับความดันโลหิตครั้งที่ 1 (BP1), ระดับคอเลสเตอรอลในเลือดครั้งที่ 2 (วัดห่างจากครั้งแรก 12 ปี) (CHOL2) และการวินิจฉัยของโรคหัวใจแบบขาดเลือด (IHDDX) เกี่ยวพันกับสถานะภาพคนไข้ที่เป็นโรคหัวใจ อย่างมีนัยสำคัญทางสถิติที่ 0.05 ตัวแบบโพรบิตที่ประมาณได้ดังใน (1)

$$\phi^{-1} [P(x)] = -3.8550 + 0.0673AGE + 0.3537BP1 + 0.6191 CHOL2(1) - 0.4078CHOL2(2) - 0.00913 IHDDX(1) + 0.7495IHDDX(2) - 0.9427 IHDDX(3) \quad (1)$$

เมื่อ ϕ แทนฟังก์ชันความหนาแน่นปกติมาตรฐาน (standard normal density function) ดังนั้นความน่าจะเป็นที่คนไข้จะเป็นโรคหัวใจ เมื่อมีสถานะภาพและผลการตรวจจากปัจจัยเสี่ยงต่าง ๆ สามารถแปลงจาก (1) ให้มีรูปแบบดังนี้

$$P(Y=1 | x) = P(x) = F(-3.8550 + 0.0673AGE + 0.3537BP1 + 0.6191 CHOL2(1) - 0.4078 CHOL2(2) - 0.00913IHDDX(1) + 0.7495IHDDX(2) - 0.9427 IHDDX(3))$$

ตารางที่ 3 ค่าประมาณพารามิเตอร์และการทดสอบพารามิเตอร์ ตัวแบบโพรบิต

พารามิเตอร์ (β)	DF	$\hat{\beta}$	S.E. ($\hat{\beta}$)	Wald	P-value
Intercept	1	-3.8550	0.7266	28.1471	<0.0001
AGE	1	0.0673	0.0137	24.1107	<0.0001
BP1	1	0.3537	0.1610	4.8269	0.0280
CHOL2(1)	1	0.6191	0.2424	6.5216	0.0107
CHOL2(2)	1	-0.4078	0.1990	4.1982	0.0405
IHDDX(1)	1	-0.00913	0.2310	0.0016	0.9685
IHDDX(2)	1	0.7495	0.2642	8.0450	0.0046
IHDDX(3)	1	-0.9427	0.4154	5.1494	0.0233

จากตารางที่ 3 พบว่าปัจจัยเสี่ยงหรือตัวแปรอธิบายที่เลือกเข้าสมการมี 4 ตัวคือ อายุ (AGE), ระดับความดันโลหิตครั้งที่ 1 (BP1), ระดับคอเลสเตอรอลในเลือดครั้งที่ 2 (วัดห่างจากครั้งแรก 12 ปี) (CHOL2) และการถูกวินิจฉัยของโรคหัวใจแบบขาดเลือด (IHDDX) โดยพบว่า ระดับปัจจัยที่ส่งผลต่อความน่าจะเป็นของระดับการเป็น

โรคหัวใจ คือ AGE, BP, CHOL2(1), CHOL2(2), IHDDX(2) และ IHDDX(3) ซึ่งให้ P-values < 0.05 นั่นคือ 0.0001, 0.0280, 0.0107, 0.0405, 0.0046 และ 0.0233 ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ 0.05 ดังนั้นจากตารางที่ 2-3 ให้ผลลัพธ์ที่แสดงปัจจัยเสี่ยงต่าง ๆ ที่ส่งผลต่อระดับการเป็นโรคหัวใจในเชิงสถิติและตัวแบบมีความเหมาะสมในการนำไปใช้พยากรณ์ด้วยตัวแบบโพรบิตต่อไป

ตัวอย่างการพยากรณ์ด้วยตัวแบบโพรบิต เช่น สมมติถ้าต้องการพยากรณ์คน ๆ หนึ่งที่มีอายุ 66 ปีและถูกวินิจฉัยว่าเป็นโรคหัวใจแบบขาดเลือดที่มีอาการกล้ามเนื้อหัวใจตายว่าจะมีความน่าจะเป็นที่คน ๆ นั้นที่มีอายุ 66 ปีและถูกวินิจฉัยว่าเป็นโรคหัวใจแบบขาดเลือดที่มีอาการกล้ามเนื้อหัวใจตาย จะเสี่ยงต่อการเสียชีวิตด้วยโรคหัวใจ โดยการกำหนดให้ AGE = 66 และ IHDDX(2) = 1 ส่วนตัวแปรอื่น ๆ มีค่าคงตัวเป็น 0 สามารถคำนวณความน่าจะเป็นของระดับการเป็นโรคหัวใจที่จะเสี่ยงต่อการเสียชีวิต คือ

$$P(x) = F(1.3363) = 0.7919 \text{ (79.19\%)}$$

หมายความว่า ความน่าจะเป็นของระดับการเป็นโรคหัวใจที่จะเสี่ยงต่อการเสียชีวิตจะเท่ากับ 79.19% ถ้าคน ๆ นั้นมีอายุ 66 ปีและถูกวินิจฉัยว่าเป็นโรคหัวใจแบบขาดเลือดที่มีอาการกล้ามเนื้อหัวใจตายในขณะนั้น เมื่อปัจจัยอื่น ๆ คงตัว

กรณีที่ 2 การวิเคราะห์ข้อมูลของตัวแปรตอบสนอง ภายใต้ตัวแบบลอจิต

ตารางที่ 4 ความเกี่ยวพันของตัวแปรตอบสนองกับปัจจัยเสี่ยง ตัวแบบลอจิต

ปัจจัยเสี่ยง (risk factors)	DF	Wald	P-value
AGE_G	1	22.5123	<0.0001
BP1	1	4.4675	0.0345
CHOL2	2	6.1601	0.0460
IHDDX	3	11.5291	0.0092

จากตารางที่ 4 ปัจจัยเสี่ยงที่ถูกเลือกเข้าตัวแบบลอจิตคือ อายุ (AGE_G), ระดับความดันโลหิตครั้งที่ 1 (BP1), ระดับคอเลสเตอรอลในเลือดครั้งที่ 2 (CHOL2) และการถูกวินิจฉัยของโรคหัวใจแบบขาดเลือด (IHDDX) โดยส่งผลต่อความน่าจะเป็นของระดับการเป็นโรคหัวใจอย่างมีนัยสำคัญทางสถิติที่ 0.05 (P-values = 0.0001, 0.0345, 0.0460, และ 0.0092 ตามลำดับ)

ตารางที่ 5 ค่าประมาณพารามิเตอร์และการทดสอบพารามิเตอร์ ตัวแบบลอจิต

พารามิเตอร์ (β)	DF	$\hat{\beta}$	S.E. ($\hat{\beta}$)	Wald	P-value
Intercept	1	0.4711	0.4457	1.1173	0.2905
AGE_G(1)	1	1.1671	0.2460	22.5123	<0.0001
BP1(1)	1	-0.6264	0.2964	4.4675	0.0345
CHOL2(1)	1	-0.9739	0.3983	5.9783	0.0145
CHOL2(2)	1	0.7143	0.3517	4.1261	0.0422
IHDDX(1)	1	0.0933	0.4179	0.0498	0.8234
IHDDX(2)	1	-1.1386	0.4627	6.0558	0.0139
IHDDX(3)	1	2.0480	0.8405	5.9376	0.0148

จากตารางที่ 5 พบว่าค่าประมาณระดับของพารามิเตอร์ของตัวแบบลอจิสต์มี 4 ตัวเช่นเดียวกันกับตัวแบบโพรบิต คือ ระดับความดันโลหิตครั้งที่ 1 (BP1), ระดับคอเลสเตอรอลในเลือดครั้งที่ 2 (CHOL2), การถูกวินิจฉัยของโรคหัวใจแบบขาดเลือด (IHDDX) และอายุ (AGE_G) และพบว่า ระดับปัจจัยเสี่ยงที่ส่งผลต่อความน่าจะเป็นของระดับการเป็นโรคหัวใจอย่างมีนัยสำคัญทางสถิติที่ 0.05 คือ AGE_G(1), BP1, CHOL2(1), CHOL2 (2), IHDDX(2) และ IHDDX(3) ซึ่งให้ P-values = 0.0345, 0.0145, 0.0422, 0.0139, 0.0148 และ <0.0001 ตามลำดับ และสามารถเขียนตัวแบบลอจิสต์ ดังนี้

$$\log\left(\frac{P(\mathbf{x})}{1 - P(\mathbf{x})}\right) = 0.4711 - 0.6264BP1(1) - 0.9739CHOL2(1) + 0.7143 CHOL2 (2) + 0.0933IHDDX(1) - 1.1386 IHDDX(2) + 2.0480 IHDDX(3) + 1.1671 AGE_G(1) \quad (2)$$

จากตัวแบบลอจิสต์สามารถวิเคราะห์ในเทอมออดส์ (odds) ดังนี้

$$P(Y=1 | \mathbf{x}) = P(\mathbf{x}) = \frac{\exp(\hat{\eta})}{1 + \exp(\hat{\eta})}$$

ดังนั้น

$$\text{odds} = \frac{P(\mathbf{x})}{1 - P(\mathbf{x})} = \exp(\hat{\eta})$$

โดยที่ $\hat{\eta} = 0.4711 - 0.6264BP1(1) - 0.9739CHOL2(1) + 0.7143 CHOL2 (2) + 0.0933IHDDX(1) - 1.1386 IHDDX(2) + 2.0480 IHDDX(3) + 1.1671 AGE_G(1)$

จากออดส์สามารถนำไปสู่การตีความหมายในเทอมของอัตราส่วนออดส์ (odds ratio) เช่นสำหรับปัจจัยอายุ พบว่า ความน่าจะเป็นของระดับการเป็นโรคหัวใจที่จะเสี่ยงต่อการเสียชีวิตของคนไข้กลุ่มที่มีอายุน้อยกว่าที่เป็นกลุ่มฐาน เมื่อปัจจัยอื่น ๆ คงตัว $\exp(1.1671) = 3.21266$ เท่า ของคนไข้กลุ่มที่มีอายุน้อยกว่าที่เป็นกลุ่มฐาน เมื่อปัจจัยอื่น ๆ คงตัว

ส่วนการตีความหมายของอัตราส่วนออดส์ภายใต้ตัวแปรอธิบายเชิงกลุ่มปัจจัยอื่น ๆ นั้นสามารถทำได้ในทำนองเดียวกันนี้ ซึ่งพบว่าตัวแปร AGE และ IHDDX มีอิทธิพลสูงกว่าตัวอื่น ๆ เช่นเดียวกันกับที่พบภายใต้ตัวแบบโพรบิตอย่างมีนัยสำคัญทางสถิติ

ตัวอย่างการพยากรณ์ด้วยตัวแบบลอจิสต์ เช่น สมมติถ้าต้องการพยากรณ์คน ๆ หนึ่งที่มีระดับคอเลสเตอรอลในเลือดครั้งที่ 2 สูง และถูกวินิจฉัยว่าเป็นโรคหัวใจแบบขาดเลือดที่มีอาการหัวใจขาดเลือดชั่วขณะว่าจะมีค่าความน่าจะเป็นที่คน ๆ นั้นจะเสี่ยงต่อการเสียชีวิต ที่มีระดับคอเลสเตอรอลในเลือดครั้งที่ 2 สูง และถูกวินิจฉัยว่าเป็นโรคหัวใจแบบขาดเลือดที่มีอาการหัวใจขาดเลือดชั่วขณะนั้น คำนวณโดยกำหนดให้ CHOL2(2) = 1 และ IHDDX(3) = 1 ส่วนตัวแปรอื่น ๆ มีค่าเป็น 0 ซึ่งคำนวณได้ค่าของ $\hat{\eta} = 3.2334$

ค่าความน่าจะเป็นที่คน ๆ นั้นจะเสี่ยงต่อการเสียชีวิต คำนวณได้ดังนี้

$$P(Y=1 | \mathbf{x}) = P(\mathbf{x}) = \frac{\exp(3.2334)}{1 + \exp(3.2334)} = 0.9620 \text{ (96.20\%)}$$

หมายความว่า ความน่าจะเป็นของระดับการเป็นโรคหัวใจที่จะเสี่ยงต่อการเสียชีวิตจะเท่ากับ 96.20% ถ้าคน ๆ นั้น มีผลการตรวจเกี่ยวกับระดับคอเลสเตอรอลในเลือดครั้งที่ 2 สูง และถูกวินิจฉัยว่าเป็นโรคหัวใจแบบขาดเลือดที่มี อาการหัวใจขาดเลือดชั่วคราว เมื่อปัจจัยอื่น ๆ คงตัว

นอกจากนี้การตีความหมายของผลลัพธ์ในเชิงการเปรียบเทียบความน่าจะเป็นที่สนใจโดยใช้ Odds Ratio ระหว่างค่าความน่าจะเป็นที่คน ๆ นั้นจะเสี่ยงต่อการเสียชีวิตเทียบกับความน่าจะเป็นที่คน ๆ นั้นจะไม่เสียชีวิต และสามารถคำนวณอัตราส่วน Odds Ratio เพิ่มเติมได้ทำนองเดียวกันกับข้างต้นต่อไป เช่น

$$\text{odds} = \frac{P(x)}{1 - P(x)} = \exp(\hat{\ln} = \exp(3.2334) = 25.3658$$

หมายความว่า อัตราส่วนระหว่างค่าความน่าจะเป็นที่คน ๆ นั้นจะเสี่ยงต่อการเสียชีวิตจะมีแนวโน้มเป็น 25.3658 เท่าของค่าความน่าจะเป็นที่คน ๆ นั้นจะไม่เสียชีวิต เมื่อเขามีระดับคอเลสเตอรอลในเลือดครั้งที่ 2 สูง และถูก วินิจฉัยว่าเป็นโรคหัวใจแบบขาดเลือดที่มีอาการหัวใจขาดเลือดชั่วคราว เมื่อปัจจัยอื่น ๆ คงตัว

กรณีที่ 3 การวิเคราะห์ข้อมูล ภายใต้ตัวแบบลอจิสติกแบบกลุ่มฐาน

ตารางที่ 6 ความเกี่ยวพันของตัวแปรตอบสนองกับปัจจัยเสี่ยงในภาพรวม ตัวแบบลอจิสติกแบบกลุ่มฐาน

ปัจจัยเสี่ยง (risk factors)	DF	Wald	P-value
BP1	3	16.6550	0.0008
AGE_G	3	25.0340	<0.0001

จากตารางที่ 6 พบว่า ปัจจัยเสี่ยงหรือตัวแปรอธิบายที่ถูกคัดเลือกด้วยวิธีการตัวแบบขั้นตอน ได้แก่ ระดับ ความดันโลหิตครั้งที่ 1 (BP1) และอายุ (AGE_G) ที่ส่งผลต่อความน่าจะเป็นของระดับการเป็นโรคหัวใจอย่างมี นัยสำคัญทางสถิติที่ 0.05 ส่วนค่าประมาณของพารามิเตอร์พร้อมด้วยการทดสอบของตัวแบบลอจิสติกแบบกลุ่มฐาน แสดงในตารางที่ 7

ตารางที่ 7 ค่าประมาณพารามิเตอร์และการทดสอบพารามิเตอร์ ตัวแบบลอจิสติกแบบกลุ่มฐาน

พารามิเตอร์ (β)	ระดับ	DF	$\hat{\beta}$	S.E. ($\hat{\beta}$)	Wald	P-value
Intercept	1	1	2.2587	0.5897	14.6693	0.0001
Intercept	2	1	1.0950	0.6305	3.0156	0.0825
Intercept	3	1	1.8321	0.6060	9.1392	0.0025
BP1(1)	1	1	-0.1527	0.5668	0.0725	0.7877
BP1(1)	2	1	-1.1184	0.6010	3.4627	0.0628
BP1(1)	3	1	0.2646	0.5977	0.1959	0.6580
AGE_G(1)	1	1	0.6751	0.4575	2.1777	0.1400
AGE_G(1)	2	1	0.3053	0.5296	0.3324	0.1400
AGE_G(1)	3	1	-0.4797	0.4465	1.1540	0.2827

จากตารางที่ 7 พบว่าตัวแปรอธิบายหรือปัจจัยเสี่ยงจำนวน 2 ปัจจัยที่เลือกเข้าสมการ คือ ระดับความดันโลหิตครั้งที่ 1 (BP1) และอายุ (AGE_G(1)) ที่ส่งผลต่อความน่าจะเป็นของระดับการเป็นโรคหัวใจอย่างมีนัยสำคัญทางสถิติที่ 0.05 และเนื่องจากตัวแปรตอบสนองที่ศึกษานี้มี 4 กลุ่ม ดังนั้นจะสามารถเขียนตัวแบบลอจิสติกเทียบกลุ่มฐานโดยเทียบแต่ละกลุ่มกับกลุ่มที่ 4 ที่เป็นกลุ่มฐานได้ 3 สมการคือ (3) - (5) ดังนี้

$$\log\left(\frac{P_1(x)}{P_4(x)}\right) = 2.2587 - 0.1527BP1(1) + 0.6751AGE_G(1) \quad (3)$$

$$\log\left(\frac{P_2(x)}{P_4(x)}\right) = 1.0950 - 1.1184BP1(1) + 0.3053AGE_G(1) \quad (4)$$

$$\log\left(\frac{P_3(x)}{P_4(x)}\right) = 1.8321 + 0.2646BP1(1) - 0.4797AGE_G(1) \quad (5)$$

ส่วนลอจิสติกเทียบกลุ่มฐานอื่น ๆ เช่นระหว่างกลุ่มที่ 1 กับ กลุ่มที่ 2 คือ $\log\left(\frac{P_1(x)}{P_2(x)}\right)$ สามารถคำนวณได้จากเซตของสมการข้างต้น (3) - (4) นั่นคือความน่าจะเป็นของเมื่อคนที่ไม่เป็นโรคหัวใจและยังมีชีวิตอยู่ เทียบกับความน่าจะเป็นคนที่เป็โรคหัวใจและยังมีชีวิตอยู่ ได้ดังสมการ (6) ต่อไปนี้

$$\log\left(\frac{P_1(x)}{P_2(x)}\right) = \log\left(\frac{P_1(x)}{P_4(x)}\right) - \log\left(\frac{P_2(x)}{P_4(x)}\right)$$

จะได้ว่า
$$\log\left(\frac{P_1(x)}{P_2(x)}\right) = 1.1637 + 0.9657BP1(1) + 0.3698AGE_G(1) \quad (6)$$

ตัวอย่างการพยากรณ์ด้วยตัวแบบลอจิสติกเทียบกลุ่มฐาน เช่น จาก (6)

สมมติถ้าต้องการพยากรณ์คน ๆ หนึ่งที่มีความดันโลหิตปกติและมีอายุน้อยกว่า 55 ปี ต้องการหาความน่าจะเป็น ซึ่งโดยทั่วไปทำในรูปเชิงเปรียบเทียบความน่าจะเป็นที่สนใจกับความน่าจะเป็นของกลุ่มฐานอีกกลุ่มหนึ่ง ๆ ที่สนใจด้วย เช่น ความน่าจะเป็นของเมื่อคนที่ไม่เป็นโรคหัวใจและยังมีชีวิตอยู่ เทียบกับความน่าจะเป็นคนที่เป็โรคหัวใจและยังมีชีวิตอยู่ สามารถคำนวณได้จากสมการ (6) ดังนี้

$$\log\left(\frac{P_1(x)}{P_2(x)}\right) = 1.1637 + 0.9657(1) + 0.3698(1) = 2.4992$$

ดังนั้น
$$\frac{P_1(x)}{P_2(x)} = \exp(2.4992) = 12.1728$$

หมายความว่า ความน่าจะเป็นที่คน ๆ นั้นจะไม่เป็นโรคหัวใจและยังมีชีวิตอยู่เป็น 12.1728 เท่าของความน่าจะเป็นที่คน ๆ นั้นจะเป็นโรคหัวใจและยังมีชีวิตอยู่ เมื่อปัจจัยอื่น ๆ คงตัว

กล่าวอีกนัยหนึ่งคือ อัตราส่วนระหว่างความน่าจะเป็นที่คน ๆ นั้นจะไม่เป็นโรคหัวใจและยังมีชีวิตอยู่มีแนวโน้มเป็น 12.1728 เท่าของความน่าจะเป็นที่คน ๆ นั้นจะเป็นโรคหัวใจและยังมีชีวิตอยู่นั่นเอง

นอกจากนี้การพยากรณ์ความน่าจะเป็นเชิงเปรียบเทียบกับกลุ่มฐานสำหรับผู้อื่น ๆ หรือระดับของตัวแปรตอบสนองผู้อื่น ๆ สามารถทำได้ในทำนองเดียวกันนี้ โดยอาศัยตัวแบบลอจิสติกเทียบกับกลุ่มฐานจากเซตของสมการ (3) - (5) รวมทั้งการคำนวณอัตราส่วนออดส์ที่สนใจต่าง ๆ สามารถทำได้ในทำนองเดียวกันดังในตัวแบบลอจิสติกภายใต้ปัจจัยเสี่ยงต่าง ๆ ที่สนใจและพบว่าส่งผลต่อตัวแปรตอบสนองอย่างมีนัยสำคัญทางสถิติ

วิจารณ์ผลการวิจัย (Discussion)

จากตัวแบบโพรบิตและตัวแบบลอจิสติก กรณีตัวแปรตอบสนองจำแนกประเภท 2 กลุ่ม พบว่า ตัวแบบโพรบิตให้ค่าสถิติ Somers'D, Gamma, Tau-a และ C โดยมีค่าของความเกี่ยวพันเท่ากับ 0.669, 0.670, 0.299 และ 0.834 ตามลำดับ และตัวแบบลอจิสติกให้ค่าสถิติ Somers'D, Gamma, Tau-a และ C โดยมีค่าของความเกี่ยวพันเท่ากับ 0.620, 0.660, 0.277 และ 0.810 ตามลำดับ จะเห็นได้ว่า ตัวแบบโพรบิตให้ค่าของตัวสถิติดังกล่าวใกล้เคียงกับของตัวแบบลอจิสติก ซึ่งต่างเป็นตัวแบบสมมาตรและมีความคล้ายคลึงกัน และเมื่อทดลอง plots ส่วนเหลือและการทดสอบความเหมาะสมของทั้งสองตัวแบบพบว่ามีความสอดคล้องกันและต่างเป็นตัวแบบที่เหมาะสมกับข้อมูลอย่างมีนัยสำคัญทางสถิติที่ 0.05

ส่วนตัวแบบลอจิสติกเทียบกับกลุ่มฐาน แม้ว่า P-values จากปัจจัยต่าง ๆ อาจสูงกว่าหรือใกล้เคียงค่าระดับนัยสำคัญทางสถิติที่ 0.05 แต่ค่าจุดตัดส่วนใหญ่ก็มีนัยสำคัญทางสถิติที่ 0.05 และการทดสอบตัวแบบลอจิสติกเทียบกับกลุ่มฐานในภาพรวมด้วยตัวสถิติ Wald นั้น พบว่ามีความเหมาะสม (ตารางที่ 6) รวมทั้งตัวสถิติ อัตราส่วนภาชนะน่าจะเป็น (P-value < 0.0001) หรือจากค่าดีเวียนซ์ (deviance) ที่แสดงไว้ในส่วนสรุปผลของแต่ละตัวแบบ ดังนั้นตัวแบบลอจิสติกเทียบกับกลุ่มฐานมีภาวะสารูปดีและเหมาะสมกับข้อมูลและสามารถใช้วิเคราะห์ข้อมูลต่อไป อนึ่ง การวิเคราะห์ข้อมูลภายใต้ตัวแบบเชิงเส้นน้อยทั่วไปนั้นมีหลายตัวแบบ กรณีศึกษาจากข้อมูลชุดเดียวสามารถตรวจสอบความเหมาะสมของแต่ละตัวแบบก่อน แล้วนำไปสู่การหาปัจจัยเสี่ยงที่ส่งผลต่อความน่าจะเป็นของระดับการเป็นโรคหัวใจอย่างมีนัยสำคัญทางสถิติ สำหรับการวิจัยครั้งต่อไปอาจยืนยันและเพิ่มประสิทธิภาพของตัวแบบนี้ให้มากขึ้นได้จากการเพิ่มจำนวนปัจจัยและจำนวนตัวอย่างที่นำมาศึกษา และอาจวิเคราะห์ข้อมูลเชิงเปรียบเทียบจากหลายแหล่ง ส่วนการศึกษาเพื่อเปรียบเทียบประสิทธิภาพระหว่างตัวแบบรวมทั้งการใช้ตัวแบบที่อาจมีทั้งอิทธิพลหลักและอิทธิพลร่วมควรใช้การจำลองแบบแทน เพื่อนำไปประยุกต์กับข้อมูลจริงต่อไป

สรุปผลการวิจัย (Conclusion)

กรณีที่ 1 การวิเคราะห์ข้อมูลของตัวแปรตอบสนอง ตัวแบบโพรบิต

ตัวแบบโพรบิตที่เหมาะสมกับกรณีนี้ อย่างมีนัยสำคัญทางสถิติที่ 0.05 คือ

$$\phi^{-1} [P(x)] = -3.8550 + 0.0673AGE + 0.3537BP1 + 0.6191 CHOL2(1) - 0.4078 CHOL2(2) \\ - 0.00913 IHDDX(1) + 0.7495IHDDX(2) - 0.9427 IHDDX(3)$$

จากตัวแบบโพรบิตที่ได้จากการวิเคราะห์ข้อมูล พบว่า ปัจจัยเสี่ยงที่ถูกคัดเลือกเข้าตัวแบบ ได้แก่ อายุ (AGE), ระดับความดันโลหิตครั้งที่ 1 (Blood pressure: BP1), ระดับคอเลสเตอรอลในเลือดครั้งที่ 2 (Cholesterol : CHOL2) และการถูกวินิจฉัยของโรคหัวใจแบบขาดเลือด (IHDDX) อย่างมีนัยสำคัญทางสถิติที่ 0.05 จากค่าสถิติ

อัตราส่วนภาวะน่าจะเป็นมีค่าเท่ากับ 56.4776 ค่าสถิติสกอ์ (Score) มีค่าเท่ากับ 46.7527 และค่าสถิติวาลด์ มีค่าเท่ากับ 37.4385 เนื่องจากค่าของ P-values มีค่าน้อยกว่าระดับนัยสำคัญ 0.05 หมายความว่าค่าประมาณพารามิเตอร์ β ของตัวแบบโพรบิตมีค่าไม่เท่ากับศูนย์ กล่าวอีกนัยหนึ่งคือ ตัวแบบที่ได้จากวิธีการดังกล่าวมีความเหมาะสมที่จะใช้ในการพยากรณ์ ภายใต้ระดับนัยสำคัญทางสถิติ 0.05 ซึ่งสอดคล้องกับตัวสถิติดีไวเยนซ์ ซึ่งมีค่าเท่ากับ 153.5720 และตัวสถิติเพียร์สัน มีค่าเท่ากับ 186.7135 และเนื่องจาก P-value จากทั้งสองค่ามากกว่า 0.05 ซึ่งหมายความว่า ตัวแบบมีภาวะสารูปดีหรือตัวแบบมีความเหมาะสมที่จะใช้ในการพยากรณ์ ภายใต้ระดับนัยสำคัญทางสถิติ 0.05

กรณีที่ 2 การวิเคราะห์ข้อมูลของตัวแปรตอบสนอง ตัวแบบลอจิต

ตัวแบบลอจิตที่เหมาะสมกับกรณีนี้อย่างมีนัยสำคัญทางสถิติที่ 0.05 คือ

$$\log\left(\frac{P(\mathbf{x})}{1 - P(\mathbf{x})}\right) = 0.4711 - 0.6264BP1(1) - 0.9739CHOL2(1) + 0.7143 CHOL2 (2) + 0.0933IHDDX(1) \\ - 1.1386 IHDDX(2) + 2.0480 IHDDX(3) + 1.1671 AGE_G(1)$$

จากตัวแบบลอจิตที่ได้จากการวิเคราะห์ข้อมูล พบว่า ปัจจัยเสี่ยงที่ถูกคัดเลือกเข้าตัวแบบ ได้แก่ ระดับความดันโลหิตครั้งที่ 1 (BP1), ระดับคอเลสเตอรอลในเลือดครั้งที่ 2 (CHOL2), การถูกวินิจฉัยของโรคหัวใจแบบขาดเลือด (IHDDX) และอายุ (AGE_G) อย่างมีนัยสำคัญทางสถิติที่ 0.05 จากตัวสถิติอัตราส่วนภาวะน่าจะเป็นมีค่าเท่ากับ 52.9252 ค่าสถิติสกอ์ มีค่าเท่ากับ 47.1652 และค่าสถิติวาลด์ มีค่าเท่ากับ 32.6528 เนื่องจากค่าของ P-values มีค่าน้อยกว่าระดับนัยสำคัญ 0.05 แสดงว่าค่าประมาณพารามิเตอร์ของตัวแบบลอจิตมีค่าไม่เท่ากับศูนย์ กล่าวอีกนัยหนึ่งคือ ตัวแบบที่ได้จากวิธีการดังกล่าวมีความเหมาะสมที่จะใช้ในการพยากรณ์ ภายใต้ระดับนัยสำคัญทางสถิติ 0.05 ซึ่งสอดคล้องกับตัวสถิติดีไวเยนซ์ มีค่าเท่ากับ 136.6699 และตัวสถิติเพียร์สัน มีค่าเท่ากับ 146.8985 เนื่องจาก P-values มีค่ามากกว่า 0.05 ซึ่งแสดงว่า ตัวแบบมีภาวะสารูปดีหรือตัวแบบมีความเหมาะสมกับข้อมูลที่จะนำไปใช้ในการพยากรณ์ ภายใต้ระดับนัยสำคัญทางสถิติ 0.05

กรณีที่ 3 การวิเคราะห์ข้อมูลของตัวแปรตอบสนอง ตัวแบบลอจิตเทียบกลุ่มฐาน

เซตของตัวแบบลอจิตเทียบกลุ่มฐานที่เหมาะสมกับกรณีนี้อย่างมีนัยสำคัญทางสถิติที่ 0.05 เช่น

$$\log\left(\frac{P_1(\mathbf{x})}{P_4(\mathbf{x})}\right) = 2.2587 - 0.1527BP1(1) + 0.6751AGE_G(1)$$

$$\log\left(\frac{P_2(\mathbf{x})}{P_4(\mathbf{x})}\right) = 1.0950 - 1.1184BP1(1) + 0.3053AGE_G(1)$$

$$\log\left(\frac{P_3(\mathbf{x})}{P_4(\mathbf{x})}\right) = 1.8321 + 0.2646BP1(1) - 0.4797AGE_G(1)$$

$$\log\left(\frac{P_1(\mathbf{x})}{P_2(\mathbf{x})}\right) = 1.1637 + 0.9657BP1(1) + 0.3698AGE_G(1)$$

และตัวแบบเทียบกลุ่มฐานคู่อื่น ๆ ที่อาจสนใจ สามารถทำได้ทำนองเดียวกันกับตัวแบบเทียบกลุ่มฐานคือ

$$\log\left(\frac{P_1(\mathbf{x})}{P_2(\mathbf{x})}\right)$$

จากตัวแบบลอจิตเทียบกลุ่มฐานที่ได้จากการวิเคราะห์ข้อมูล พบว่า ปัจจัยเสี่ยงที่ถูกเลือกเข้าตัวแบบ ได้แก่ ระดับความดันโลหิตครั้งที่ 1 (BP1) และอายุ (AGE_G) อย่างมีนัยสำคัญทางสถิติที่ 0.05 (ตารางที่ 6) จากค่าสถิติอัตราส่วนภาวะน่าจะเป็น มีค่าเท่ากับ 45.4538 ค่าสถิติสกอร์ มีค่าเท่ากับ 49.9336 และค่าสถิติวาลด์ มีค่าเท่ากับ 39.9706 เนื่องจากค่าของ P-values มีค่าน้อยกว่าระดับนัยสำคัญ 0.05 หมายความว่าค่าประมาณพารามิเตอร์ในภาพรวมของตัวแบบลอจิตเทียบกลุ่มฐานมีค่าแตกต่างจากศูนย์ กล่าวอีกนัยหนึ่งคือ ตัวแบบที่ได้จากวิธีการดังกล่าวมีความเหมาะสมที่จะใช้ในการพยากรณ์ ภายใต้ระดับนัยสำคัญทางสถิติ 0.05

การพยากรณ์ความน่าจะเป็นของตัวแปรตอบสนองของกลุ่มต่าง ๆ กันของทุกตัวแบบ จะขึ้นอยู่กับผู้ศึกษาว่า จะสนใจทำนายกลุ่มหรือระดับใดของตัวแปรตอบสนอง และผู้ช่วยมีอาการหรือมีลักษณะเฉพาะและ/หรือมีผลการตรวจพบอย่างหนึ่งอย่างใดหรือไม่ โดยสามารถเลือกตัวแบบใดตัวแบบหนึ่งหรือหลายตัวแบบข้างต้น เพื่อนำไปประกอบการตัดสินใจในทางปฏิบัติจริงตามความเหมาะสมที่พบปัจจัยเสี่ยงหนึ่ง ๆ ที่ส่งผลกระทบต่อตัวแปรตอบสนองดังกล่าว

เอกสารอ้างอิง (References)

- จุฑามาศ สสิโรจน์ ฆฎายุรัตน์ ชิตชิตกุล ดารารัตน์ ศรีเดช และ วีรานันท์ พงศาภักดี (2557). การใช้ตัวแบบเชิงเส้นน้อยทั่วไปในการวิเคราะห์ปัจจัยเสี่ยงต่อการเสียชีวิตของทารก ที่พิจารณาจากน้ำหนักแรกเกิดของทารก วารสารวิทยาศาสตร์ มข. ฉบับที่ 2 ปีที่ 42 (เมษายน-มิถุนายน 2557): 434-448.
- วีรานันท์ พงศาภักดี. (2555). การวิเคราะห์ข้อมูลจำแนกประเภท: ทฤษฎีและการประยุกต์ ด้วย GLIM, SPSS, SAS และ MTB พิมพ์ครั้งที่ 3. นครปฐม: โครงการตำราคณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร โรงพิมพ์มหาวิทยาลัยศิลปากร. หน้า. 165-196.
- <http://www.saimhelath.net> ค้นเมื่อวันที่ 18 มกราคม 2557
- <http://www.thaiheart.org> ค้นเมื่อวันที่ 18 มกราคม 2557
- <http://www.wikipedia.org> ค้นเมื่อวันที่ 18 มกราคม 2557
- Affii A.A. and Azen S.P. (1979). Statistical Analysis: A Computer Oriented Approach. 2nd ed. New York: Academic Press. pp. 345-428.
- Agresti, A. (2002). Categorical Data Analysis. 2nd ed. New York: John Wiley & Sons. pp. 267-311.
- Agresti, A (2007). Introduction to Categorical Data Analysis. New York: John Wiley & Sons. pp. 173-179.
- Agresti, A. (2013) Categorical Data Analysis. 3rd ed. New York: John Wiley & Sons. pp. 341, 344, 346, 350, 352.
- Bliss, C.I. (1935). The Calculation of the Dosage-mortality Curve. Ann. Appl. Biol. 22: 134-167.
- Hosmer, D, W. and Lemeshow, S. (1989). Applied Logistic Regression. New York: John Wiley & Sons. pp. 142-143.
- Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized Linear Models. J. Roy. Statist. Soc. Ser A135(3): 370-384.
- SAS (Enterprise Guide 5.1) (2013) Statistical Analysis System Software, Enterprise Guide 5.1 License of Department of statistics, Faculty of Science, Silpakorn University.